ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas, Juleen Lam, Patrice Sutton, Paula I. Johnson, Dylan S. Atchley, Saunak Sen, Karen A. Robinson, Daniel A. Axelrad, and Tracey J. Woodruff

http://dx.doi.org/10.1289/ehp.1307177
Received: 4 June 2013
Accepted: 23 April 2014
Advance Publication: 25 June 2014

NIH National Institute of Environmental Health Sciences

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas,[1] Juleen Lam,[1] Patrice Sutton,[2] Paula I. Johnson,[2] Dylan S. Atchley,[2] Saunak Sen,[3] Karen A. Robinson,[4] Daniel A. Axelrad,[5] and Tracey J. Woodruff[2]

[1]Oak Ridge Institute for Science and Education (ORISE) Post-doctoral Fellow with the U.S. Environmental Protection Agency, Office of Policy, National Center for Environmental Economics, Washington, DC, USA; [2]University of California San Francisco, Program on Reproductive Health and the Environment, Oakland, California, USA; [3]University of California San Francisco, Department of Epidemiology and Biostatistics, San Francisco, California, USA; [4]Johns Hopkins University, Departments of Medicine, Epidemiology and Health Policy & Management, Baltimore, Maryland, USA; [5]U.S. Environmental Protection Agency, Office of Policy, National Center for Environmental Economics, Washington, DC, USA

Address correspondence to Erica Koustas, U.S. Environmental Protection Agency, 1200 Pennsylvania Ave. NW (1809T), Washington, DC 20460 USA. Telephone: (202) 566-1829. E-mail: Koustas.Erica@epa.gov

Running title: Systematic review of PFOA and fetal growth

Acknowledgments: We thank Kate Guyton (U.S. Environmental Protection Agency) for her contribution to developing the evidence evaluation methodologies and assistance with identifying relevant toxicological databases. We thank Andy Rooney (National Toxicology Program) for his toxicological expertise and for assistance with developing animal risk of bias

criteria. We also thank Kristina Thayer (National Toxicology Program), Lisa Bero (University of California, San Francisco), and Lauren Zeise (California Environmental Protection Agency) for assistance with developing risk of bias evaluation criteria, meta-analysis methodologies, and evidence evaluation methodologies. Malcolm Macleod (University of Edinburgh) provided critical assistance in developing the animal-specific systematic review methodology. We thank Hanna Vesterinen (University of California, San Francisco) for her thoughtful comments on this manuscript. Christopher Lau (U.S. Environmental Protection Agency) served as a PFOA subject matter expert. We thank Tara Horvath (University of California, San Francisco) for training and assistance with developing our search strategy. Finally, we acknowledge Barbara Abbott, Julie Boberg, John Butenhoff, Suzanne Fenton, An Hagenaars, Jiayin Dai, Jamie DeWitt, Christopher Lau, Gerald Kennedy, Sean Kennedy, Jason O'Brien, Natalia Onishchenko, Adi Pinkas, Sally White, Joseph Yanai, and their laboratories for their kind and timely response to requests for additional information.

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. The research was supported in part by appointments to the Internship/Research Participation Program at the National Center for Environmental Economics, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. Dr. Woodruff is also a member of the Philip R. Lee Institute for Health Policy Studies, UCSF.

Financial support for development of the PFOA case study was through grants from New York Community Trust and the U.S. Environmental Protection Agency through a contract with Abt Associates (GAIA-0-6-UCSF 17288).

Support over the past 5 years for the development and dissemination of the Navigation Guide methodology was provided by: Clarence Heller Foundation, Passport Foundation, Forsythia Foundation, the Johnson Family Foundation, the Heinz Endowments, the Fred Gellert Foundation, Rose Foundation, Kaiser Permanente, New York Community Trust, University of California, San Francisco Institute for Health Policy Studies, Planned Parenthood Federation of America, National Institute for Environmental Health Sciences (ES018135 and ESO22841) and U.S. Environmental Protection Agency STAR (RD83467801 and RD83543301).

Competing financial interests: The authors declare no competing financial interests.

Abstract

Background: In contrast to current methods of "expert-based narrative review", the Navigation Guide is a systematic and transparent method for synthesizing environmental health research from multiple evidence streams. The Navigation Guide was developed to effectively and efficiently translate the available scientific evidence into timely prevention-oriented action.

Objectives: Apply the Navigation Guide systematic review method to answer the question: "Does fetal developmental exposure to perfluorooctanoic acid (PFOA) or its salts affect fetal growth in animals," and rate the strength of the experimental animal evidence.

Methods: We conducted a comprehensive search of the literature, applied pre-specified criteria to the search results to identify relevant studies, extracted data from studies, obtained additional information from study authors, conducted meta-analyses, and rated the overall quality and the strength of the evidence.

Results: Twenty-one studies met the inclusion criteria. From the meta-analysis of eight mouse gavage datasets, we estimated that exposure of pregnant mice to increasing concentrations of PFOA was associated with a decrease in mean pup birth weight of -0.023g (95% CI: -0.029, -0.016) per 1-unit increase in dose (mg/kg BW/day). The evidence, consisting of 15 mammalian and 6 non-mammalian studies, was rated as 'moderate' and 'low' quality, respectively.

Conclusion: Based on this first application of the Navigation Guide methodology, we found 'sufficient' evidence that fetal developmental exposure to PFOA reduces fetal growth in animals.

Introduction

Background

In clinical research, systematic reviews have played a transformative role as a transparent, robust method for synthesizing the available evidence for incorporation into more efficient guidelines and recommendations related to medical interventions. But while systematic review methodology has been developed and tested in the clinical sciences for making evidence-based decisions for medical interventions, the methods are not fully transferable to environmental health science, largely because of their primary application to randomized controlled clinical trials, which are, for primarily ethical reasons, unavailable in environmental health. The Navigation Guide was developed to bridge this gap between clinical and environmental health sciences. The methodology provides the capacity to systematically and transparently evaluate the quality and strength of evidence from both human and non-human streams of evidence about the relationship between the environment and reproductive and developmental health (Woodruff and Sutton 2011; Woodruff and Sutton 2014).

To test and refine the Navigation Guide systematic review methodology, we applied it to the evaluation of experimental animal evidence for the effects of exposure to the environmental contaminant perfluorooctanoic acid (PFOA) on fetal growth. The results of applying the method to the human evidence and integrating the animal and human data into an overarching strength of evidence rating are presented elsewhere (Johnson et al. 2014; Lam et al. 2014).

Rationale for selecting PFOA

Environmental exposures to the industrial chemical PFOA are widespread and PFOA has been detected in the blood of over 95% of the U.S. population (ATSDR 2009; CDC 2009; Kato et al.

2011; US EPA 2009) and in blood samples throughout the world (Kannan et al. 2004; US EPA 2009). Voluntary efforts by eight major manufacturers of PFOA to eliminate global emissions and product content by the end of 2015 are ongoing, and significant progress has been made for both U.S. and non-U.S. operations (US EPA 2008, 2012b, 2013a). However, PFOA can remain in the environment, and with a half-life in humans of approximately 3.5 years (Olsen et al. 2007), the chemical will persist in people for years to come (US EPA 2012a, 2013b).

Fetal exposure to PFOA may be widespread as the chemical is ubiquitous in blood of pregnant women, women of child-bearing age, and in cord blood (Apelberg et al. 2007a; Calafat et al. 2007; Fei et al. 2007; Midasch et al. 2007; Mondal et al. 2012; Monroy et al. 2008; Woodruff et al. 2011). The association between PFOA and fetal growth reported in individual human studies has been inconsistent, with some reporting statistically significant associations between prenatal exposure to PFOA and restricted fetal growth (Apelberg et al. 2007b; Fei et al. 2007, 2008) and others reporting no or non-statistically significant associations (Hamm et al. 2010; Monroy et al. 2008; Washino et al. 2009). The animal literature also includes reports of inconsistent associations between PFOA and fetal growth, including findings of reduced birth weight following prenatal exposure to PFOA in rodent studies (Butenhoff et al. 2004; Hines et al. 2009; Lau et al. 2004; Lau et al. 2006).

Ubiquitous exposure to a chemical that lacks evidence of non-toxicity is a potential public health concern; moreover, PFOA has been associated with adverse impacts on the quality and duration of the gestation period--one of the most important indicators of an infant's health and survival (Gluckman and Hanson 2006; Institute of Medicine 2007). Due to the potential concern for an adverse developmental health outcome of public health importance and the availability of data, we selected PFOA to test and refine the Navigation Guide method.

Methods

A priori, we assembled a review team to include experts in the fields of risk assessment, environmental health, epidemiology, biology, systematic review, and toxicology to develop a protocol that covered the first three steps of the Navigation Guide systematic review method: (1) Specify the study question; (2) Select the evidence; and (3) Rate the quality and strength of the evidence (Koustas et al. 2013) available at: http://prhe.ucsf.edu/prhe/navigationguide.html and summarized below. Each of the steps of the Navigation Guide method described below involves application of standardized and transparent documentation including expert judgment. Additional information regarding the Navigation Guide methodology can be found elsewhere (Woodruff and Sutton 2014).

Step 1. Specify the study question

Our objective was to answer the question: "Does fetal developmental exposure to PFOA or its salts affect fetal growth in animals?" "PICO" (Participants, Interventions, Comparator, Outcomes) is an aid used to formulate an answerable question in a systematic review, and to provide more specific information about the scope of the review (O'Connor et al. 2011). Because we were evaluating environmental exposures, we used the acronym "PECO" i.e., "**P**articipants", "**E**xposure," "**C**omparator" and "**O**utcomes."

Participants: Animals from non-human species studied during reproductive/developmental time period (before and/or during pregnancy for females or during development for embryos).

Exposure: One or more oral, subcutaneous or other treatment(s) of any dosage with PFOA, CAS# 335-67-1, or its salts during the time before pregnancy and/or during pregnancy for females or directly to embryos.

Comparator: Experimental animals receiving different doses of PFOA or vehicle-only treatment.

Outcomes: For mammalian species: fetal weight near term (for example, embryonic day 18 for mice and embryonic day 21 for rat) or at birth; and/or other measures of size near term or at birth, such as length. For non-mammalian species: weight and/or other measures of size in late stages of embryonic development.

Step 2. Select the evidence

Search methods

Our search was developed by analyzing the Medical Subject Headings (MeSH) and other terms from the title and abstract text of a group of seven papers known to us, judged to be relevant to our study question, and which represented different journals and years of publication (Abbott et al. 2007; Butenhoff et al. 2004; Hines et al. 2009; Lau et al. 2006; Staples et al. 1984; White et al. 2007; White et al. 2009). A list of common and unique terms was compiled and incorporated into a search strategy to address the exposure (PFOA) and outcomes of interest (reproductive/developmental toxicity), as defined in the PECO statement (see Supplemental Material, Tables S1-S2). To develop search terms to retrieve experimental animal studies, we adapted a search filter developed by Hooijmans et al. (Hooijmans et al. 2010b).

We searched PubMed and Web of Science on February 3, 2012. We searched 35 toxicological databases using PFOA terms between January 23 and February 6, 2012. Our search was not limited by language or publication date. We hand searched the reference list of all included studies and searched for publications citing the included studies. We also consulted with a subject matter expert (Christopher Lau, US EPA).

Data collection and management

We imported or manually entered all retrieved records into EndNote (X4) reference management software, and each record was assigned a source identification number, which was used to track individual studies throughout the course of the review. Two authors (E.K. and J.L.) independently screened the titles and abstracts of each record retrieved to identify those meeting our inclusion criteria using DistillerSR (Evidence Partners; available at: http://www.systematic-review.net). We developed inclusion and exclusion criteria based on our PECO statement. All studies that compared experimental animals exposed to one or more doses of PFOA during reproductive or developmental periods to untreated control experimental animals were eligible for inclusion. We excluded studies if one or more of the following criteria were met: article did not contain original data (i.e., review article); study subjects were not animals; PFOA was not administered to study subjects; PFOA was not administered during reproductive/developmental time period. Two authors (E.K. and J.L.) assessed the full-text of studies that could not be excluded based on the title and abstract screening. Potentially relevant non-English articles were translated to determine eligibility. To provide quality control, a third author (P.I.J.) screened the title and abstract of five percent of the search results and five percent or five articles, whichever was greater, of search results eligible for full text. We considered studies that described more than one experiment or outcome measure as separate datasets.

Two authors (E.K. and J.L.) independently extracted data relating to study characteristics and outcome measures from all included articles into a Microsoft Access (2007) database. The list of extracted study characteristics was based on a compilation of previously published checklists and criteria (Guyatt et al. 2011; Higgins and Deeks (editors) 2011; Hooijmans et al. 2010a; Kilkenny et al. 2010).

One author (E.K.) performed data entry of the raw outcome data using Microsoft Excel (2007) and a second author (D.S.A.) verified all values. We contacted study authors when additional information was required for performing statistical analysis and/or analysis of the full dataset. For example, we requested numerical estimates associated with figures presented in published articles, numbers of animals allocated to various test groups, and raw data values. In some cases, fetal growth data were not presented in the published study because the outcome was not of primary interest to the study authors. If there was a reason to believe the study authors may have measured fetal growth, we contacted them to obtain any data they may have collected during the course of the study. We also contacted all study authors to inform them of our systematic review, and to verify values used in both the meta-analysis and analysis of the full dataset.

Statistical analyses

Two authors (E.K. and J.L.) assessed study characteristics from all included articles for comparability (i.e., study features and biological heterogeneity) to determine which studies were suitable for meta-analysis. We consulted experts in the field of PFOA toxicity, toxicological study design, or human/animal toxicity reviews to develop these characteristics and their associated heterogeneity concerns *a priori*. For example, we considered PFOA clearance rate differences between female mice (approximately 17 days) and rats (2-4 hours) as a potential biological heterogeneity concern (Lau et al. 2007).

From the assessment of specified characteristics, we determined that only a subset of data was combinable in a meta-analysis. This subset of seven studies (eight datasets) had the following characteristics:

- Species: mouse

- Route of exposure: gavage
- Method of outcome measurement: weight
- Time point of outcome measurement: at or soon after birth

We used the mean pup body weight at birth (and standard error) from each of the eight datasets, for all doses below 5 mg/kg body weight (BW)/day. The dose was limited to focus on effects at lower tested doses and to minimize adverse impacts from responses at higher doses (such as litter loss) on the overall estimate. We used a two-step modeling approach. In the first step, we analyzed each dataset separately using a linear mixed effects model, and obtained a slope estimate of the dose-response effect (and associated standard error). In the second step, we combined the slope and standard error estimate from each dataset using a random effects model.

The result was an estimate of the overall mean change in body weight per offspring for a 1-unit increase in mg/kg BW per day dose, accounting for within- and between-study variability. We used the programming environment R version 2.13.1 (R Development Core Team; available at: http://www.R-project.org/) and its standard packages. We used the R package "metafor" (Viechtbauer, 2010) for conducting our random effects meta-analysis.

In order to visually assess the possibility of publication bias in a meta-analysis, we considered producing a funnel plot of the estimated effects. However, tests for funnel plot asymmetry are not recommended when there are fewer than ten studies because test power is usually too low to distinguish chance from real asymmetry (Sterne et al. 2011a). As our meta-analysis was limited to seven studies (eight datasets), we did not produce a funnel plot.

Statistical heterogeneity assessment

We sought to assess whether differences in estimated effect sizes among studies were consistent with random variation versus non-random heterogeneity among the studies. We estimated the between study variance component, and tested the null hypothesis that the between study variability was absent using Cochran's Q statistic. The test statistic follows a Chi-squared distribution with n-1 degrees of freedom, where n is the number of studies. We considered a p-value of 0.05 or less statistically significant. We also calculated the I^2 statistic, which estimates the percentage of variation across studies due to heterogeneity rather than chance (Higgins et al. 2003), and used the Cochrane Collaboration's guidelines to interpret the statistic, where a value of greater than 50% may indicate substantial heterogeneity (Deeks et al. 2011). To assess the overall impact of existing study heterogeneity on the meta-analysis, we considered the magnitude/direction of effect estimates, the I^2 statistic, and the p-value from the Cochran's Q test.

Sensitivity analysis

We performed sensitivity analyses using subgroup analyses based on characteristics described above that were used to determine comparability across studies for the meta-analysis. To evaluate the influence of each individual study on the main meta-analysis results, and assist in identifying any study characteristics that might be influential in the final results, we performed a sensitivity analysis by removing one dataset at a time from the meta-analysis.

Analysis of the full dataset

We analyzed all included animal studies identified via our search and exclusion/inclusion assessment to assess the totality of all available animal evidence. This was done to maximize use

of all data, in addition to those determined appropriate to combine in the meta-analysis. To assess results from the full dataset, we calculated percentage change in outcome (weight or length) compared to the control group for each tested dose group for each of the datasets and used these values to create scatter plots. Two of the non-mammalian studies reported outcome measurements at multiple time points during larval development (Spachmo and Arukwe 2012; Wang et al. 2010). We selected the outcome measurement reported at the latest time point during the larval stage, based on justification that this allowed for consideration of maximal larval growth. For each study, we used the mean and standard error estimates reported by authors to calculate a 95% confidence interval for the difference in means comparing each treatment group to the control group. We interpreted a 95% confidence interval that overlapped zero as indicating no statistically significant difference between the mean weight in that treatment group with the mean weight in the control group.

Step 3. Rate the quality and strength of the evidence

To rate the evidence, we 1) determined risk of bias for individual studies based on seven domains; 2) rated the overall quality across all studies in the body of evidence based on five factors, including risk of bias; and 3) rated the overall strength of evidence across all studies in the body of evidence based on four considerations, including quality of body of evidence (Figure 1).

Assessment of risk of bias

Two authors (E.K. and J.L.) assessed risk of bias defined as "methodological characteristics of a study that can introduce a systematic error in the magnitude or direction of the results" (Higgins and Altman (editors) 2011) for included studies based on seven risk of bias domains using

modified terminology and concepts in the Cochrane Collaboration's Tool for Assessing Risk of Bias. Informed by empirical data from meta-analyses conducted on pharmacological treatments (Roseman et al. 2011), we considered funding source and reported conflicts of interest to be potential sources of bias. We did not ask study authors for additional information to inform our risk of bias determinations. However, if study authors mentioned study design details in their responses to our requests for data, we considered the information while evaluating risk of bias. See Table 1 for a summary of the risk of bias domains assessed for each included study and Supplemental Material, "Instructions for making risk of bias determinations".

Rate the quality and strength of the body of evidence

Upon completion of the data analysis, each of the nine review authors compared the results of the systematic review to the Navigation Guide factors and considerations for rating the quality and strength of the non-human evidence. The Navigation Guide rating method (Woodruff and Sutton 2011) was applied according to explicit written directions (Koustas et al. 2013). Due to fundamental biological differences between mammalian and non-mammalian model systems, we evaluated the mammalian and non-mammalian studies as separate bodies of evidence.

The possible ratings for the overall quality of the body of evidence were 'high', 'moderate', and 'low'. These quality ratings were determined by assigning an initial rating according to the type of study, and then downgrading the rating if factors that decrease the quality level of the studies were present. The initial quality rating assigned to both the mammalian and non-mammalian bodies of evidence was 'high', comparable to the rating assigned to human experimental studies, i.e., randomized controlled trials (RCTs), in systematic review methods used in the clinical sciences. An initial 'high' quality rating for experimental animal studies was supported by the level of study control exercised in such studies and the limited heterogeneity within an

experimental animal study population. This is also consistent with Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines for clinical evidence that consider randomization a key determinant of 'high' grade (Guyatt et al. 2011). Upgrades to the quality rating for experimental animal data were not considered because the initial quality level was 'high.'

The overall body of evidence was evaluated for downgrading based on the presence of five factors (Figure 1):

(1) Risk of Bias Across Studies: Study limitations – a substantial risk of bias across body of evidence;

(2) Indirectness: Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, and/or outcome). *A priori*, we decided not to downgrade experimental animal studies for indirectness, as studies suggest that humans are as or more sensitive to chemical exposures than animals, strengthening the applicability of findings from experimental animal studies to human health outcomes (Kimmel et al. 1984; US EPA 1996). However, in applying GRADE principles to the Navigation Guide, evidence would be rated down if it is determined that the animal model is biologically inappropriate for the health outcome under study.

(3) Inconsistency: Widely different estimates of effect (heterogeneity or variability in results);

(4) Imprecision: Studies had few participants and few events (wide confidence intervals); and

(5) Publication Bias: Studies missing from the body of evidence, resulting in an overestimate or underestimate of true effects from exposure.

According to GRADE, these five factors address nearly all issues that bear on the quality of evidence (Balshem et al. 2011). Each of the nine review authors reviewed the body of evidence and applied their expert judgment to independently and transparently grade the quality of evidence based on the presence of the five objective factors using detailed instructions (Koustas et al. 2013). Possible ratings were 0 (no change from 'high' quality), -1 (1 level downgrade to 'moderate' quality) or – 2 (2 level downgrade to 'low' quality). Consistent with GRADE's approach to evaluating risk of bias across studies (Guyatt et al. 2011), authors were instructed to be conservative in making judgments to downgrade the evidence for all factors (i.e. high confidence of substantial concerns with the body of evidence before rating down). Authors reviewed the body of evidence as a way to initiate the group discussion and gather all perspectives for consideration. After independently evaluating the quality of the evidence, all authors discussed their evaluations. The discussion between co-authors was extensive, iterative, and carried out over several meetings until a consensus was reached. These collective decisions did not involve a "majority vote" or other tallying of perspectives. It was specified *a priori* that discrepancies between review authors' that could not be resolved through consensus would be resolved by the senior author (TW). However, for this case study, review authors were able to agree on a collective consensus for each rating and the arbiter was not necessary. The rationale for each collective decision on each of the five factors was recorded.

In systematic reviews in the clinical sciences, rating the quality of evidence is the final step, because only one stream of evidence is considered in a decision. However, as our purpose was to ultimately integrate the strength of multiple streams of evidence used in environmental health decision-making, i.e., toxicology and epidemiology, leading to a concise "bottom line" statement about a chemical's toxicity that brings all of the relevant evidence to bear, the Navigation Guide

systematic review method specifies an additional step --- moving from quality of evidence to strength of evidence.

We rated the overall strength of the evidence based on a combination of four considerations: (1) Quality of body of evidence; (2) Direction of effect estimates; (3) Confidence in effect estimates (likelihood that a new study would change our conclusion); and (4) Other compelling attributes of the data that may influence certainty (Figure 1). The results of rating the strength of the non-human evidence were compared to the definitions specified in the Navigation Guide for 'sufficient' evidence of toxicity; 'limited' evidence of toxicity; 'inadequate' evidence of toxicity; or 'evidence of lack of toxicity' (Table 2), which were based on criteria in use by the International Agency for Research on Cancer (IARC) and U.S. EPA (IARC 2006; US EPA 1991, 1996). The procedure for rating the strength of the evidence was similar to rating the quality of evidence: all review authors independently evaluated the strength of the evidence according to the same four considerations, then compared their evaluations, resolved any discrepancies through discussion, and recorded the rationale for every collective decision.

Results

Included studies

We identified 2,049 unique records (see Supplemental Material, Table S3 for the total number of hits retrieved from each database) of which 1,982 were excluded through title and abstract screening, and 46 articles excluded during full-text review, resulting in 21 studies describing 32 datasets for inclusion in the review (Figure 2). A summary of mammalian and non-mammalian study characteristics are provided in Tables 3 and 4, respectively. Detailed characteristics of each mammalian and non-mammalian study are provided in Supplemental Material, Tables S4-18 and Tables S19-S24, respectively. Various details of outcome data and study design characteristics

necessary for data analysis were missing from all 21 articles. In some cases, published articles did not include details needed for our analysis, such as numerical outcome measurements or data on fetal growth if this was not a primary outcome of interest for study authors. In other cases, basic information, such as allocation numbers or the number of animals weighed to obtain given outcome values, was missing. Our efforts to contact study authors resulted in obtaining additional data for 18 of 21 included studies, along with raw data in many instances (see Supplemental Material, Tables S4-S24).

Populations

Of 21 studies, 15 were conducted on mammalian species (11 mouse and four rat) and six studies were conducted on non-mammalian species (three chicken, one fruit fly, one zebrafish, and one salmon) (Tables 3 and 4).

Mammalian exposures

For all 15 mammalian studies, pregnant female dams were exposed to PFOA and fetal growth was measured in the resulting progeny (Table 3). The primary route of exposure was oral gavage (13 studies), but some studies also evaluated exposures via inhalation, food and water. The majority (12) of mammalian studies exposed dams to the ammonium salt form of PFOA (CAS#3825-26-1), one study exposed dams to the free acid form (CAS# 335-67-1), and two studies did not specify the form used for exposure. The dose range tested varied widely across studies, ranging from 0.01-100 mg/kg BW/day. Inhalation study doses ranged between 0.1-25 mg/m^3. The number of PFOA doses administered per study ranged from one to six. While dams in all studies were exposed to PFOA at some point during their pregnancy, the window of exposure varied across studies from a single gavage exposure on a single day of pregnancy to exposure prior to conception that continued throughout pregnancy.

Mammalian comparators

Eleven gavage studies used water as a vehicle control and two used corn oil (Table 3). The inhalation study utilized three control groups: in-house air only, in-house air pair-fed to 10 mg/m^3 dose group, and in-house air pair-fed to 25 mg/m^3 dose group. The PFOA-treated food study used food applied with ethanol as a control and the PFOA-treated water study used non-treated water as a control. Besides PFOA exposure, all control groups were treated similarly to dose groups for each dataset.

Mammalian outcomes

Body weight was used as the outcome measure for all 15 mammalian studies (Table 3). Because pregnant dams were exposed to PFOA for all mammalian studies, the litter was used as the statistical unit and the total number analyzed across studies ranged from eight to 183 litters.

The time point of weight measurement varied between fetal time points near term, typically gestation day 18 (GD18) for mice and GD21 for rats, to at or near the time of birth, typically postnatal day 0 (PND0) to PND2. The methods used to monitor parturition varied widely across birth weight studies, from constant monitoring to daily cage checks. PND1 was defined as either the day of birth or the day after birth.

The method of weight measurement varied across studies as well, from measuring offspring individually, grouped by litter or by sex, to measuring a subset of offspring from each litter. Offspring survival was statistically significantly reduced (based on the alpha level specified by study authors, generally <0.05 or ≤ 0.05) at exposure to doses above 5 mg/kg BW/day in five studies and one study did not provide statistics or comment on litter sizes at birth.

Mammalian risk of bias assessment

Based on our risk of bias assessment, we concluded that the majority of studies had *probably high* risk of bias for 'allocation concealment' and 'blinding', and *probably low* risk of bias for 'incomplete outcome data' and 'selective reporting'. Ratings for 'sequence generation' and 'conflict of interest' were mixed across studies, and ranged from *low* to *high* risk of bias. All studies had *low* risk of bias for the 'other bias' domain (Figure 3A and 3B). See Supplemental Material, Tables S25-S39 for details on the risk of bias results for each mammalian study.

Non-mammalian exposures

Developing embryos were directly exposed to PFOA in all six non-mammalian studies (Table 4). Routes of administration varied based on test species: injection of PFOA solution into eggs for chicken studies; immersion of eggs in PFOA solution for zebrafish and salmon studies; PFOA-treated food for fruit fly studies. One study exposed organisms to the ammonium salt form of PFOA (CAS#3825-26-1), two studies exposed organisms to the free acid form of PFOA (CAS#335-67-1), and three studies did not specify the form of PFOA. The dose ranges across studies varied based on animal species tested: chicken (0.01-10 mg/kg egg); zebrafish (15-250 mg/L water); fruit fly (100-500 μM in food); salmon (100 μg/L water). The number of PFOA doses administered per study ranged from one to eight.

All non-mammalian studies exposed embryos during development and the time period of exposure varied based on species. For the chicken studies, a single injection of PFOA was administered to eggs on incubation day 0; for zebrafish studies, eggs were exposed 60 minutes after spawning to 120 hours post fertilization (hpf); for salmon studies, eggs and larvae were exposed to PFOA-containing water for 48 days; for fruit fly studies, female flies were allowed to

lay eggs for two hours in vials with PFOA-containing food, and eggs were allowed to hatch and develop through 110 hours after egg laying (ael) or to white pupae stage, depending on dataset.

Non-mammalian comparators

Chicken studies used saline, dimethyl sulfoxide (DMSO), or sunflower oil as vehicle controls and some studies included an un-injected control (Table 4). The zebrafish study used water as a vehicle control, the fruit fly study used untreated food as a vehicle control, and the salmon study used water with carrier solvent (methanol) as a vehicle control. Besides PFOA exposure, all control groups were treated similarly to dose groups for each dataset.

Non-mammalian outcomes

Relevant outcome measures varied across non-mammalian studies and included length, weight, and larval volume (calculated from measurements of length and diameter) (Table 4). Because embryos were directly exposed to PFOA in the non-mammalian model systems, the embryo was used as the unit of statistical analysis and the total number of embryos analyzed across studies varied between 37 and 378.

The time points of outcome measurement varied from shortly before time of hatching, shortly after hatching, and multiple time points during larval development.

PFOA exposure delayed hatching/larval emergence in the zebrafish and fruit fly studies and induced mortality in the zebrafish study and in one chicken study. Pipping success (i.e., when a chick breaks its shell) and developmental stage at embryo death were unaffected by PFOA exposure in one chicken study, while in a second chicken study, embryonic mortality was increased, but hatchling mortality and hatching success were not affected. The salmon study did not provide details on larval survival rates.

Non-mammalian risk of bias assessment

Based on our risk of bias assessment, we found that the majority of studies had *probably high* risk of bias for 'sequence generation', 'allocation concealment', and 'blinding', and *probably low* risk of bias for 'selective reporting'. Ratings for 'incomplete outcome data' were mixed across studies, and ranged from *low* to *high* risk of bias. Finally, all studies had *probably low* or *low* risk of bias for 'conflict of interest' and *low* risk of bias for the 'other bias' domain (Figure 3A and 3B). See Supplemental Material, Tables S40-S45 on details of the risk of bias results for each non-mammalian study.

Impact of PFOA on fetal growth

Analysis

Across the eight datasets determined to be combinable in the meta-analysis, gavage exposure of pregnant mice to increasing concentrations of PFOA was associated with a decrease in birth weight. The combined estimate from the meta-analysis was a change in mean pup birth weight of -0.023g [95% CI -0.029, -0.016] per 1-unit increase in dose (mg/kg BW/day) (Figure 4). The I^2 test statistic was calculated to be 0%, indicating no observed heterogeneity between studies that could not be explained by chance; this conclusion was further supported by the Q statistic, which produced a non-significant p-value of 0.73.

We found from the sensitivity analysis, when removing one dataset at a time, that there were relatively small changes in the effect estimate with a maximum of 9% change in the meta-analysis estimate (from -0.023 to -0.021) seen when removing the White 2011 dataset (White et al. 2011) (data not shown). Figure 4 shows that this particular study resulted in the largest estimate of decreased birth weight among those studies weighted more heavily in the meta-

analysis (indicated by the larger size of the mean symbol), so it is not surprising that the removal of this study would have the largest effect on the meta-analysis estimate, and in particular shifting it to a smaller estimate of decreased birth weight. Although the Abbott 2007 dataset (Abbott et al. 2007) had the largest effect estimate, removing the dataset had little effect on the meta-analysis due to its small weight. The sensitivity analysis further demonstrated that the 95% confidence intervals were also minimally affected, and consistently did not include 0.

We created separate scatter plots to summarize all mammalian study data for near-term, fetal weight measurements (Figure 5A) and for birth weight measurements (Figure 5B). The dose-response data for the nine studies not included in the meta-analysis showed mixed results, generally with lower doses showing increased weight compared to the control group (mostly non-significant) and higher doses showing decreased weight (both statistically significant and not) (Figure 5B). The 95% confidence intervals for the mean difference comparing birth weight in the treatment versus control group for each study are presented in Supplemental Material, Tables S46 and S47.

We also created scatter plots to summarize non-mammalian study data, separately for weight measurements (Figure 6A) and for length measurements (Figure 6B). A qualitative evaluation of dose-response data showed mostly non-statistically significant increases in body weight, even at the highest tested doses. The length data show mixed results, with two studies demonstrating statistically significant decreases in length and the other two studies showing non-significant increases in length. The 95% confidence intervals for the mean difference comparing birth weight in the treatment versus control group for each study are presented in Supplemental Material, Tables S48 and S49.

Quality of evidence

We downgraded the overall quality rating of the mammalian evidence from 'high' to 'moderate' based on the 'risk of bias across studies' criterion, as the majority of studies were deemed to have *probably high* risk of bias for the 'allocation concealment' and 'blinding' domains. Our ratings and rationales for the overall quality of mammalian evidence are presented in Table 5.

We downgraded the overall quality rating of the non-mammalian evidence from 'high' to 'low' due to: (1) 'risk of bias across studies,' given that the majority of studies were deemed to have *probably high* risk of bias for the 'sequence generation', 'allocation concealment', and 'blinding' domains; and (2) 'indirectness' as, for the purpose of this case study, we did not have a rationale or evidence to support that all the non-mammalian species and their corresponding routes of exposure were directly applicable model systems for evaluating human fetal growth. Our ratings and rationales for the overall quality of non-mammalian evidence are presented in Table 6.

Strength of evidence rating

We excluded the non-mammalian data from the final strength of evidence rating. Our rationale was that the non-mammalian evidence was judged to be of 'low' quality for the purposes of addressing our study question, and we had higher quality direct evidence on which to base a decision. Our strength of the evidence considerations were as follows:

- Quality of body of evidence: 'Moderate'
- Direction of effect estimates: Decreasing birth weight with increasing exposure to PFOA
- Confidence in effect estimates: Confidence based on the consistency of the results and overlapping confidence intervals
- Other compelling attributes of the data that may influence certainty: None

We compared these considerations to the definitions in Table 2 and concluded that the animal evidence is 'sufficient' to conclude that exposure to PFOA or its salts adversely affect fetal growth in animals.

Discussion

Animal evidence for PFOA and fetal growth

Based on this first application of the Navigation Guide systematic review methodology, we found 'sufficient' evidence that fetal developmental exposure to PFOA or its salts reduces fetal growth in animals. Our finding that the data were 'sufficient' was based on 'moderate' quality mammalian evidence, reduction in mean offspring birth weight from dams exposed to increasing concentrations of PFOA during pregnancy, and our confidence in the effect based on the consistency of the results and overlapping confidence intervals. Analysis of the scatter plots of the studies excluded from the meta-analysis supported that the majority of these studies also found consistently small reductions in measures of fetal growth following maternal exposure to PFOA.

From the meta-analysis of eight mouse gavage datasets, we estimated that exposure of pregnant mice to increasing concentrations of PFOA was associated with a decrease in mean pup birth weight of -0.023g (95% CI: -0.029, -0.016) per 1-unit increase in dose (mg/kg BW/day). To assess the biological significance of this estimate, we pooled birth weight measurements from each of the eight control groups to estimate an overall mean birth weight of 1.57g for the pups in control groups. A 0.023g decrease in body weight is equivalent to an approximate 1.46% decrease in average body weight per 1-unit increase in PFOA dose. Thus, for example, according to this model, a dose of 10 mg/kg BW/day PFOA to pregnant dams is estimated to cause approximately a 15% decrease in the litter's average birth weight.

To address the heterogeneity of the available evidence, we limited the meta-analysis to data from mouse studies. The rationale for this decision was based in part on findings from pharmacokinetic studies documenting that the rate of elimination for PFOA is much faster for female rats, as compared to other mammalian species, including humans (Lau et al. 2007). Many of the studies included in our meta-analysis cited rate of elimination differences as a supporting reason for using mouse model systems. However, responses between mouse model systems may differ as well; evidence suggests that responses to PFOA may vary based on the mouse strain tested. One study noted that the 129S1/SvlmJ strain was more sensitive to PFOA exposure, as compared to the CD-1 strain (Abbott et al. 2007). We included data from the 129S1/SvlmJ strain in our meta-analysis, since, in the absence of evidence supporting which mouse strain best matches human sensitivity to PFOA, there was no evidence to support a premise that humans are less sensitive than the most sensitive mouse. This is further supported by studies of agents known to cause reproductive toxicity, for which "humans appear to be as or more sensitive than the most sensitive animal species tested" (US EPA 1996). Additionally, our sensitivity analysis found removing this study from the meta-analysis resulted in minimal changes in the meta-analysis estimate (<2%) (data not shown).

The heterogeneity of the non-mammalian animal data precluded combining these studies quantitatively. Our identification of studies among such diverse species was unexpected, and for this case study, we combined all non-mammalian species into a single body of evidence. This did not impede decision-making about toxicity of PFOA and fetal growth because more direct mammalian and human data were available. However, for other chemicals, heterogeneous indirect evidence may be the only data available on which to base a decision. This points to the need to anticipate and plan for the analysis of heterogeneous data, including if it is appropriate to

evaluate each species separately, and to determine relevance to human health *a priori* in future protocols.

Application of the Navigation Guide Systematic Review Methodology

We found the application of the method to be effective in producing a concise statement of health hazard in a systematic and transparent manner. While ultimately our review did not identify any studies relevant to our study question that were published in languages other than English, it is difficult to predict in which cases excluding non-English studies may bias a systematic review (Sterne et al. 2011b), so for future reviews we would retain this strategy. Moreover, our systematic search identified over 1900 studies that we did not find in a search we conducted at the initiation of the project using traditional non-systematic methods and our improved search strategy nearly doubled the number of studies that met our *a priori* inclusion criteria.

Despite a steep learning curve, designing and completing the search, eliminating duplicate records, screening studies, and extracting study characteristics and data took about two to three months, including time to train review authors. Contact with study authors to obtain additional information took place over the course of approximately three months. Risk of bias assessment, data analysis, and evaluation of quality and strength of evidence took approximately two to three additional months.

An inevitable limitation of this first case study was that we were simultaneously developing and applying the method. As a result, we did not anticipate or define *a priori* all the benchmarks we ultimately used for making judgments when rating the quality and strength of the evidence, and we found that our decision-making was more difficult in the absence of *a priori* definitions. To

guide our judgments when assessing quality and strength of evidence factors that had not been pre-specified, we: (1) sought an empirical basis for a judgment; (2) conducted further analysis (i.e., sensitivity); (3) relied on GRADE's principle to be conservative in the judgment of rating down; and (4) always documented the rationale for our judgment. Anticipating and defining *a priori* criteria for as many judgments as possible will improve the method; however it seems unlikely that all judgments can be anticipated. Thus, the principles we used for *post hoc* judgments will be integrated into future protocols to transparently allow for such circumstances.

Challenges in translating experimental animal evidence into improved health outcomes

In applying the Navigation Guide systematic review methodology, we found that the high prevalence of sub-optimal experimental animal study design and reporting that has been empirically documented in the preclinical literature (Bebarta et al. 2003; Landis et al. 2012; Macleod et al. 2004; Macleod et al. 2008; McPartland et al. 2007; van der Worp et al. 2007; van der Worp and Macleod 2011; Vesterinen et al. 2011) may also be prevalent in the experimental animal data that inform decision-making in environmental health. In nearly all of the studies included in our review, direct evidence to support risk of bias ratings, such as clear descriptions of randomization or blinding methods, was missing. Furthermore, many studies failed to report some of the basic data necessary for interpretation of results and incorporation into meta-analysis. For example, multiple studies failed to report data such as the number of animals included in outcome measurements (e.g., number of litters assessed, number of pups per litter, etc.), details on how offspring were weighed (e.g., individually, as a whole litter, etc.), or the time point of outcome assessment (e.g., clear definition of PND1, monitoring of parturition, etc.). In order to create scatter plots and perform a meta-analysis, we needed to contact the lead author for every study to obtain missing data. Fortunately, authors for the majority of studies responded

and many generously took the time and effort to provide raw data for inclusion in this review. Our follow-up with the authors indicated that many of these missing data were a result of deficiencies in reporting and point to the need to include contacting study authors as a step in the protocol.

These findings underscore the urgency of calls for improved experimental animal study design and reporting in the preclinical arena (Beronius et al. 2014; Krauth et al. 2013; Landis et al. 2012; van der Worp and Macleod 2011; Vesterinen et al. 2010; Vesterinen et al. 2011). To this end, a 2012 major stakeholder meeting by the U.S. National Institute of Neurological Disorders and Stroke found that at a minimum studies should report on sample-size estimation, whether and how animals were randomized, whether investigators were blind to the treatment, and the handling of data (Landis et al. 2012). It will be important for environmental health scientists and journals that publish environmental health research to help support these nascent efforts to advance the translational relevance of animal evidence into improved health outcomes (Howells and Macleod 2013; Macleod et al. 2009; van der Worp et al. 2010; Vesterinen et al. 2011).

Summary and Conclusion

This case study documents that the Navigation Guide methodology can be used to effectively apply the rigor of evidence synthesis methods in use by the clinical sciences to questions in environmental health. The Navigation Guide methodology does not eliminate the need for expert judgment, but does make clear the evidence that informs the authors' judgments and requires transparency and an explicit accounting of the judgments involved.

In addition to this review of the animal evidence, a separate systematic review was conducted evaluating the human evidence relevant to PFOA exposure and fetal growth, which resulted in a

'sufficient' evidence of toxicity rating (Johnson et al. 2014). In another paper, the strength of the evidence ratings from the non-human and human evidence were combined according to the factors specified in the Navigation Guide (Woodruff and Sutton 2011), resulting in an overall conclusion that human exposure to PFOA is 'known to be toxic' to human reproduction and development based on 'sufficient' evidence of decreased fetal growth in both human and non-human mammalian species (Lam et al. 2014). Together, these reviews demonstrate the utility of the Navigation Guide in systematically approaching a complex body of scientific evidence.

The ultimate goal of our efforts is to refine the Navigation Guide systematic review methodology across diverse streams of evidence and to support the development of recommendations for prevention in clinical and policy spheres. As has been demonstrated in the clinical field, the adoption of systematic and transparent methods to synthesize the scientific evidence in the environmental health field would speed incorporation of research into decision-making.

References

Abbott BD, Wolf CJ, Schmid JE, Das KP, Zehr RD, Helfant L, et al. 2007. Perfluorooctanoic acid induced developmental toxicity in the mouse is dependent on expression of peroxisome proliferator activated receptor-alpha. Toxicol Sci 98:571-581.

Apelberg BJ, Goldman LR, Calafat AM, Herbstman JB, Kuklenyik Z, Heidler J, et al. 2007a. Determinants of fetal exposure to polyfluoroalkyl compounds in Baltimore, Maryland. Environ Sci Technol 41:3891-3897.

Apelberg BJ, Witter FR, Herbstman JB, Calafat AM, Halden RU, Needham LL, et al. 2007b. Cord serum concentrations of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in relation to weight and size at birth. Environ Health Perspect 115:1670-1676.

ATSDR (Agency for Toxic Substances & Disease Registry). 2009. Toxicological Profile for Perfluoroalkyls. (Draft for Public Comment). Available: http://www.atsdr.cdc.gov/toxprofiles/tp200.pdf [accessed 13 February 2013].

Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. 2011. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64:401-406.

Bebarta V, Luyten D, Heard K. 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med 10:684-687.

Beronius A, Molander L, Rudén C, Hanberg, A. 2014. Facilitating the use of non-standard in vivo studies in health risk assessment of chemicals: a proposal to improve evaluation criteria and reporting. J Appl Toxicol; doi: 10.1002/jat.2991 [Online 30 January 2014].

Boberg J, Metzdorff S, Wortziger R, Axelstad M, Brokken L, Vinggaard AM, et al. 2008. Impact of diisobutyl phthalate and other PPAR agonists on steroidogenesis and plasma insulin and leptin levels in fetal rats. Toxicology 250:75-81.

Butenhoff JL, Kennedy GL, Jr., Frame SR, O'Connor JC, York RG. 2004. The reproductive toxicology of ammonium perfluorooctanoate (APFO) in the rat. Toxicology 196:95-116.

Calafat AM, Kuklenyik Z, Reidy JA, Caudill SP, Tully JS, Needham LL. 2007. Serum concentrations of 11 polyfluoroalkyl compounds in the U.S. population: data from the national health and nutrition examination survey (NHANES). Environ Sci Technol 41:2237-2242.

CDC (Centers for Disease Control and Prevention). 2009. Fourth National Report on Human Exposure to Environmental Chemicals. Atlanta, GA: Centers for Disease Control and Prevention. Available: http://www.cdc.gov/exposurereport/ [accessed 6 February 2013].

Deeks JJ, Higgins JPT, Altman DG (editors). 2011. Chapter 9:Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, (Higgins JPT, Green S, eds) Available: http://www.cochrane-handbook.org [accessed 15 April 2013].

Fei C, McLaughlin JK, Tarone RE, Olsen J. 2007. Perfluorinated chemicals and fetal growth: a study within the Danish National Birth Cohort. Environ Health Perspect 115:1677-1682.

Fei C, McLaughlin JK, Tarone RE, Olsen J. 2008. Fetal growth indicators and perfluorinated chemicals: a study in the Danish National Birth Cohort. Am J Epidemiol 168:66-72.

Fenton SE, Reiner JL, Nakayama SF, Delinsky AD, Stanko JP, Hines EP, et al. 2009. Analysis of PFOA in dosed CD-1 mice. Part 2. Disposition of PFOA in tissues and fluids from pregnant and lactating mice and their pups. Reprod Toxicol 27:365-372.

Gluckman PD, Hanson MA. 2006. The developmental origins of health and disease. Cambridge:Cambridge University Press.

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. 2011. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 64:407-415.

Hagenaars A, Vergauwen L, De Coen W, Knapen D. 2011. Structure-activity relationship assessment of four perfluorinated chemicals using a prolonged zebrafish early life stage test. Chemosphere 82:764-772.

Hamm MP, Cherry NM, Chan E, Martin JW, Burstyn I. 2010. Maternal exposure to perfluorinated acids and fetal growth. J Expo Sci Environ Epidemiol 20:589-597.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ 327:557-560.

Higgins JPT, Altman DJ (editors). 2011. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, (Higgins JPT, Green S, eds). Available: www.cochrane-handbook.org [accessed 7 February 2013].

Higgins JPT, Deeks JJ (editors). 2011. Chapter 7: Selecting studies and collecting data. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 7 February 2013].

Hinderliter PM, Mylchreest E, Gannon SA, Butenhoff JL, Kennedy GL Jr. 2005. Perfluorooctanoate: Placental and lactational transport pharmacokinetics in rats. Toxicology 211:139-148.

Hines EP, White SS, Stanko JP, Gibbs-Flournoy EA, Lau C, Fenton SE. 2009. Phenotypic dichotomy following developmental exposure to perfluorooctanoic acid (PFOA) in female CD-1 mice: Low doses induce elevated serum leptin and insulin, and overweight in mid-life. Mol Cell Endocrinol 304:97-105.

Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. 2010a. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. Altern Lab Anim 38:167-182.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2010b. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Lab Anim 44:170-175.

Howells DW, Macleod MR. 2013. Evidence-based translational medicine. Stroke 44:1466-1471.

Hu Q, Strynar MJ, DeWitt JC. 2010. Are developmentally exposed C57BL/6 mice insensitive to suppression of TDAR by PFOA? J Immunotoxicol 7:344-349.

IARC (International Agency for Research on Cancer). 2006. Preamble to the IARC Monographs. Lyon, France: International Agency for Research on Cancer. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed 7 February 2013].

Institute of Medicine. 2007. Preterm Birth: Causes, Consequences, and Prevention. Washington, DC:The National Academies Press.

Jiang Q, Lust RM, Strynar MJ, Dagnino S, DeWitt JC. 2012. Perflurooctanoic acid induces developmental cardiotoxicity in chicken embryos and hatchlings. Toxicology 293:97-106.

Johnson PI, Sutton P, Atchley D, Koustas E, Lam J, Sen S, et al. 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307893 [Online 25 June 2014].

Kannan K, Corsolini S, Falandysz J, Fillmann G, Kumar KS, Loganathan BG, et al. 2004. Perfluorooctanesulfonate and related fluorochemicals in human blood from several countries. Environ Sci Technol 38:4489-4495.

Kato K, Wong LY, Jia LT, Kuklenyik Z, Calafat AM. 2011. Trends in exposure to polyfluoroalkyl chemicals in the U.S. Population: 1999-2008. Environ Sci Technol 45:8037-8045.

Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG. 2010. Animal research: reporting in vivo experiments: the ARRIVE guidelines. Br J Pharmacol 160:1577-1579.

Kimmel CA, Holson JF, Hogue CJ, Carlo GL. 1984. Reliability of experimental studies for predicting hazards to human development. NCTR technical report for experiment 6015. Jefferson, AR:National Center for Toxicological Research.

Koustas E, Lam J, Sutton P, Johnson P, Atchley D, Sen S, et al. 2013. Applying the Navigation Guide: Case Study #1. The Impact of Developmental Exposure to Perfluorooctanoic Acid (PFOA) On Fetal Growth. A Systematic Review of the Non-Human Evidence Protocol Available: http://www.prhe.ucsf.edu/prhe/navigationguide.html [accessed 3 September 2013].

Krauth D, Woodruff TJ, Bero L. 2013. Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review. Environ Health Perspect; doi:101289/ehp1206389 [Online 14 June 2013].

Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The Navigation Guide—evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307929 [Online 25 June 2014].

Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490:187-191.

Lau C, Butenhoff JL, Rogers JM. 2004. The developmental toxicity of perfluoroalkyl acids and their derivatives. Toxicol Appl Pharmacol 198:231-241.

Lau C, Thibodeaux JR, Hanson RG, Narotsky MG, Rogers JM, Lindstrom AB, et al. 2006. Effects of perfluorooctanoic acid exposure during pregnancy in the mouse. Toxicol Sci 90:510-518.

Lau C, Anitole K, Hodes C, Lai D, Pfahles-Hutchens A, Seed J. 2007. Perfluoroalkyl acids: a review of monitoring and toxicological findings. Toxicol Sci 99:366-394.

Macleod MR, O'Collins T, Howells DW, Donnan GA. 2004. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 35:1203-1208.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. 2008. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39:2824-2829.

Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. 2009. Good laboratory practice: preventing introduction of bias at the bench. Stroke 40:e50-52.

McPartland JM, Glass M, Pertwee RG. 2007. Meta-analysis of cannabinoid ligand binding affinity and receptor distribution: interspecies differences. Br J Pharmacol 152:583-593.

Midasch O, Drexler H, Hart N, Beckmann MW, Angerer J. 2007. Transplacental exposure of neonates to perfluorooctanesulfonate and perfluorooctanoate: a pilot study. Int Arch Occup Environ Health 80:643-648.

Mondal D, Lopez-Espinosa MJ, Armstrong B, Stein CR, Fletcher T. 2012. Relationships of perfluorooctanoate and perfluorooctane sulfonate serum concentrations between mother-child pairs in a population with perfluorooctanoate exposure from drinking water. Environ Health Perspect 120:752-757.

Monroy R, Morrison K, Teo K, Atkinson S, Kubwabo C, Stewart B, et al. 2008. Serum levels of perfluoroalkyl compounds in human maternal and umbilical cord blood samples. Environ Res 108:56-62.

Mylchreest E. 2003. PFOA: Lactational and placental transport pharmacokinetic study in rats. Report prepared for 3M, St. Paul, MN. Laboratory Project ID: DuPont-13309. Reviewed in US EPA AR 226-1551. Newark, DE:Haskell Laboratory for Health and Environmental Sciences.

O'Brien JM, Crump D, Mundy LJ, Chu S, McLaren KK, Vongphachan V, et al. 2009. Pipping success and liver mRNA expression in chicken embryos exposed in ovo to C8 and C11 perfluorinated carboxylic acids and C10 perfluorinated sulfonate. Toxicol Lett 190:134-139.

O'Connor D, Green S, Higgins JPT (editors). 2011. Chapter 5: Defining the review question and developing criteria for including studies. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, (Higgins JPT, Green S, eds) Available: http://www.cochrane-handbook.org [accessed 15 April 2013].

Olsen GW, Burris JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, et al. 2007. Half-life of serum elimination of perfluorooctanesulfonate,perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. Environ Health Perspect 115:1298-1305.

Onishchenko N, Fischer C, Wan Ibrahim WN, Negri S, Spulber S, Cottica D, et al. 2011. Prenatal exposure to PFOS or PFOA alters motor function in mice in a sex-related manner. Neurotox Res 19:452-461.

Pinkas A, Slotkin TA, Brick-Turin Y, Van der Zee EA, Yanai J. 2010. Neurobehavioral teratogenicity of perfluorinated alkyls in an avian model. Neurotoxicol Teratol 32:182-186.

Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, et al. 2011. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. JAMA 305:1008-1017.

Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. 2007. Update on the methods of the U.S. Preventative Services Task Force: estimating certainty and magnitude of net benefit. Ann Intern Med 147: 871-875.

Spachmo B, Arukwe A. 2012. Endocrine and developmental effects in Atlantic salmon (Salmo salar) exposed to perfluorooctane sulfonic or perfluorooctane carboxylic acids. Aquat Toxicol 108:112-124.

Staples RE, Burgess BA, Kerns WD. 1984. The embryo-fetal toxicity and teratogenic potential of ammonium perfluorooctanoate (APFO) in the rat. Fundam Appl Toxicol 4:429-440.

Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. 2011a. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 343:d4002.

Sterne JAC, Egger M, Moher D (editors). 2011b. Chapter 10: Addressing reporting bias. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 22 April 2013].

US EPA (U.S. Environmental Protection Agency). 1991. Guidelines for Developmental Toxicity Risk Assessment. Fed Reg 56:63798-63826.

US EPA (U.S. Environmental Protection Agency). 1996. Guidelines for Reproductive Toxicity Risk Assessment. Fed Reg 61:56274-56322.

US EPA (U.S. Environmental Protection Agency). 2008. News Release: EPA Announces Substantial Decrease of PFOA. Available: http://yosemite.epa.gov/opa/admpress.nsf/68b5f2d54f3eefd28525701500517fbf/8f9dbdd044050f71852573e50064439f!OpenDocument [accessed 3 February 2013].

US EPA (U.S. Environmental Protection Agency). 2009. Long-Chain Perfluorinated Chemicals (PFCs) Action Plan. Available: http://www.epa.gov/oppt/existingchemicals/pubs/pfcs_action_plan1230_09.pdf [accessed 3 February 2013].

US EPA (U.S. Environmental Protection Agency). 2012a. Emerging Contaminants – Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic Acid (PFOA). Available: http://www.epa.gov/fedfac/pdf/emerging_contaminants_pfos_pfoa.pdf [accessed 15 April 2013].

US EPA (U.S. Environmental Protection Agency). 2012b. News Release: Industry Progressing in Voluntary Effort to Reduce Toxic Chemicals / Companies have successfully developed over 150 alternatives to long-chain perfluorinated chemicals. Available: http://yosemite.epa.gov/opa/admpress.nsf/79c090e81f0578738525781f0043619b/335d29e8d5e40864852579a0005f3760!OpenDocument [accessed 11 February 2013].

US EPA (U.S. Environmental Protection Agency). 2013a. 2012 Annual Progress Reports. Available: http://www.epa.gov/oppt/pfoa/pubs/stewardship/preports6.html [accessed 25 September 2013]

US EPA (U.S. Environmental Protection Agency). 2013b. America's Children and the Environment, 3rd Edition. Available: http://www.epa.gov/ace/publications/ACE3_2013.pdf [accessed 19 February 2013].

van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. 2007. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain 130:3063-3074.

van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. 2010. Can animal models of disease reliably inform human studies? PLoS Med 7:e1000245.

van der Worp HB, Macleod MR. 2011. Preclinical studies of human disease: time to take methodological quality seriously. J Mol Cell Cardiol 51:449-450.

Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. 2010. Improving the translational hit of experimental treatments in multiple sclerosis. Multiple Sclerosis 16:1044-1055.

Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, Dirnagl U. 2011. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the Journal of Cerebral Blood Flow and Metabolism. J Cereb Blood Flow Metab 31:1064-1072.

Viechtbauer W. 2010. Conducting meta-analyses in R with the metafor package. J Stat Softw 36:1-48.

Wang J, Li Y, Liu Y, Zhang H, Dai J. 2010. Disturbance of perfluorooctanoic acid on development and behavior in Drosophila larvae. Environ Toxicol Chem 29:2117-2122.

Washino N, Saijo Y, Sasaki S, Kato S, Ban S, Konishi K, et al. 2009. Correlations between prenatal exposure to perfluorinated chemicals and reduced fetal growth. Environ Health Perspect 117:660-667.

White SS, Calafat AM, Kuklenyik Z, Villanueva L, Zehr RD, Helfant L, et al. 2007. Gestational PFOA exposure of mice is associated with altered mammary gland development in dams and female offspring. Toxicol Sci 96:133-144.

White SS, Kato K, Jia LT, Basden BJ, Calafat AM, Hines EP, et al. 2009. Effects of perfluorooctanoic acid on mouse mammary gland development and differentiation resulting from cross-foster and restricted gestational exposures. Reprod Toxicol 27:289-298.

White SS, Stanko JP, Kato K, Calafat AM, Hines EP, Fenton SE. 2011. Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. Environ Health Perspect 119:1070-1076.

Wolf CJ, Fenton SE, Schmid JE, Calafat AM, Kuklenyik Z, Bryant XA, et al. 2007. Developmental toxicity of perfluorooctanoic acid in the CD-1 mouse after cross-foster and restricted gestational exposures. Toxicol Sci 95:462-473.

Woodruff TJ, Sutton P. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood) 30:931-937.

Woodruff TJ, Zota AR, Schwartz JM. 2011. Environmental chemicals in pregnant women in the United States: NHANES 2003-2004. Environ Health Perspect 119:878-885.

Woodruff TJ, Sutton P. 2014. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect; http://dx.doi.org/10.1289/ehp.1307175 [Online 25 June 2014].

Yahia D, El-Nasser MA, Abedel-Latif M, Tsukuba C, Yoshida M, Sato I, et al. 2010. Effects of perfluorooctanoic acid (PFOA) exposure to pregnant mice on reproduction. J Toxicol Sci 35:527-533.

York R. 2002. Oral (gavage) two-generation (one litter per generation) reproduction study of ammonium perfluorooctanoate (APFO) in rats. Report prepared for 3M, St. Paul, MN. Sponsor's Study No T-6889.6. Reviewed in US EPA AR226-1092. Horsham, PA:Argus Research.

York RG, Kennedy GL, Jr., Olsen GW, Butenhoff JL. 2010. Male reproductive system parameters in a two-generation reproduction study of ammonium perfluorooctanoate in rats and human relevance. Toxicology 271:64-72.

Table 1. Tool for assessing risk of bias.

Domain	Criteria for low risk of bias rating	Examples of factors considered by authors
Sequence generation	Study authors reported the use of a random component in the sequence generation process.	Use of a random component, such as a random number table or computer random number generator; statement by study author that animals randomly allocated.
Allocation concealment	Study authors reported that study personnel could not foresee which animals were allocated to the various experimental groups.	Use of sequentially numbered cages or animals.
Blinding	Study authors reported that personnel and outcome assessors were adequately prevented from knowledge of the allocated exposures during the study.	Use of masked identifiers in study and for outcome assessment.
Incomplete outcome data	Study authors reported when and why participants left the study.	Number of animals allocated to experimental groups reported and/or adequate follow up of dams and offspring (for mammalian studies); number of organisms allocated to experimental groups reported and/or adequate follow up of organisms following exposure (for non-mammalian studies).
Selective reporting	The study's pre-specified outcomes that are of interest in the review were reported in a pre-specified way.	Number of animals or organisms analyzed for outcomes of interest reported or study author provided additional data; study methods matched study results for outcomes of interest.
Conflict of interest	The study is free of support from a company, study author, or other entity having a financial interest in the exposures of interest in the review.	The study was funded or conducted by companies with a financial interest in PFOA; companies provided services to assist in the completion of the study, evaluate the data, or write the manuscript; the publication or report included a declaration of conflicts of interest.
Other bias	Study appears to be free of other sources of bias	Other potential sources of bias related to the specific study design

Table 2. Strength of evidence definitions for non-human studies.[a]

Strength rating	Definition
Sufficient evidence of toxicity	A positive relationship is observed between exposure and adverse outcome in multiple studies or a single appropriate study in a single species.[b] The available evidence includes results from one or more well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies.[c]
Limited evidence of toxicity	The data suggest a positive relationship between exposure and adverse outcome, but there are important limitations in the quality of the body of evidence. Confidence in the relationship is constrained by such factors as: the number, size, or quality of individual studies, or inconsistency of findings across individual studies.[c] As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
Inadequate evidence of toxicity	The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.
Evidence of lack of toxicity	Data on an adequate array of endpoints from more than one study with at least two species showed no adverse effects at doses that were minimally toxic in terms of inducing an adverse effect. Information on pharmacokinetics, mechanisms, or known properties of the chemical class may also strengthen the evidence.[d] Conclusion is limited to the species, age at exposure, and/or other conditions and levels of exposure studied, and is unlikely to be strongly affected by the results of future studies.[c]

[a]The Navigation Guide rates the quality and strength of evidence of human and non-human evidence streams separately as 'sufficient', 'limited', 'inadequate' or 'evidence of lack of toxicity' and then these two ratings are combined to produce one of five possible statements about the overall strength of the evidence of a chemical's reproductive/developmental toxicity. The methodology is adapted from the criteria used by the International Agency for Research on Cancer (IARC) to categorize the carcinogenicity of substances (IARC 2006) except as noted. [b]IARC's criteria for sufficient evidence of carcinogenicity in animals requires multiple positive results (species, studies, sexes). The Navigation Guide integrates USEPA's minimum criteria for animal data for a reproductive or developmental hazard, i.e., data demonstrating an adverse reproductive effect in a single appropriate, well-executed study in a single test species (US EPA 1996). The Navigation Guide also incorporates USEPA's "sufficient evidence category" which includes data that "collectively provide enough information to judge whether or not a reproductive hazard

exists within the context of effect as well as dose, duration, timing, and route of exposure. This category may include both human and experimental animal evidence" (US EPA 1996). The USEPA statement for developmental hazards is slightly different but includes the same relevant information regarding dose, duration, timing, etc. (US EPA 1991). [c]Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit (Sawaya et al. 2007). [d]Based on minimum data requirements according to USEPA Guidelines for Reproductive Toxicity (US EPA 1996).

Table 3. Summary of mammalian study characteristics.

Source [source ID]	Species	Time point of outcome measurement	Outcome measure	Route of exposure	Period of exposure	PFOA dose range[a]	Number of doses administered[b]	Number of litters	Reason(s) excluded from meta-analysis
Studies used in meta-analysis									
Hines et al. 2009 [260]	Mouse	Birth	Weight	Gavage	GD1-17	0.01-5	5	75	NA
White et al. 2009 [312]	Mouse	Birth	Weight	Gavage	GD8-17	5	1	8	NA
Abbott et al. 2007 [528]	Mouse	Birth	Weight	Gavage	GD1-17	0.1-1	4	58	NA
White et al. 2007 [566]	Mouse	Birth	Weight	Gavage	GD1-17, GD8-17, GD12-17	5	1	37	NA
Wolf et al. 2007 [571]	Mouse	Birth	Weight	Gavage	GD1-17	3-5	2	87	NA
Wolf et al. 2007 [571]	Mouse	Birth	Weight	Gavage	GD7-17, GD10-17, GD13-17, GD15-17	5-20	2	56	NA
Lau et al. 2006 [635][c]	Mouse	Birth	Weight	Gavage	GD1-17	1-20	5	103	NA
White et al. 2011 [3862]	Mouse	Birth	Weight	Gavage	GD1-17	1-5	2	60	NA
Studies not used in meta-analysis									
Hu et al. 2010 [68]	Mouse	Birth	Weight	Drinking water	GD6-17	0.05-1	2	30	Incomparable route of exposure
Yahia et al. 2010 [103]	Mouse	Fetal	Weight	Gavage	GD0-17	1-10	3	29	Incomparable time point of outcome measurement
Yahia et al. 2010 [103]	Mouse	Birth	Weight	Gavage	GD0-18	1-10	3	20	Time point of birth weight measurement was not specified
Fenton et al. 2009 [264]	Mouse	Fetal	Weight	Gavage	GD17	0.1-5	3	19	Incomparable time point of outcome measurement
Fenton et al. 2009 [264]	Mouse	Birth	Weight	Gavage	GD17	0.1-5	3	19	Dams were exposed for only one day of pregnancy
Lau et al. 2006 [635][c]	Mouse	Fetal	Weight	Gavage	GD1-17	1-40	6	183	Incomparable time point of outcome measurement
Hinderliter et al. 2005 [711][d]	Rat	Birth	Weight	Gavage	GD4-21	3-30	3	20	Incomparable species

Source [source ID]	Species	Time point of outcome measurement	Outcome measure	Route of exposure	Period of exposure	PFOA dose range[a]	Number of doses administered[b]	Number of litters	Reason(s) excluded from meta-analysis
Staples et al. 1984 [1871]	Rat	Fetal	Weight	Gavage	GD6-15	100	1	46	Incomparable species and time point of outcome measurement
Staples et al. 1984 [1871]	Rat	Fetal	Weight	Inhalation	GD6-15	0.1-25 mg/m^3	4[e]	103	Incomparable species, route of exposure, and time point of outcome measurement
Staples et al. 1984 [1871]	Rat	Birth	Weight	Gavage	GD6-15	100	1	21	Incomparable species
Staples et al. 1984 [1871]	Rat	Birth	Weight	Inhalation	GD6-15	0.1-25 mg/m^3	4	54	Incomparable species and route of exposure
Boberg et al. 2008 [3061]	Rat	Fetal	Weight	Gavage	GD7-20/21	20	1	11	Incomparable species and time point of outcome measurement
Onishchenko et al. 2011 [3610]	Mouse	Birth	Weight	Food	GD1-20	0.3	1	15	Incomparable route of exposure
York 2002 [5122][f]	Rat	Birth	Weight	Gavage	70 days prior to breeding through lactation	1-30	4	141	Incomparable species

(GD) = gestation day.

[a]mg/kg BW/day, unless otherwise specified; dose range is limited to those doses for which dams were analyzed. [b]Excludes control groups; 1 control group unless otherwise specified. [c]Lau 2006 study appears two times (birth weight data were included in meta-analysis; fetal data were excluded from meta-analysis). [d]Hinderliter 2005 study is a peer-reviewed publication; author provided an industry report with detailed data (Mylchreest 2003). [e]3 control groups. [f]York 2002 is an industry report; search also identified peer-reviewed journal publications describing findings from the report (Butenhoff et al. 2004; York et al. 2010), but these journal publications were excluded as duplicates since report provided raw data.

Table 4. Summary of non-mammalian study characteristics.

Source [source ID]	Species	Time point(s) of outcome measurement	Outcome measure	Route of exposure	Period of exposure	PFOA dose range	Number of doses administered[a]	Number of offspring
Hagenaars et al. 2011 [59]	Zebrafish	120 hpf (post-hatching)	Length	Egg immersion	Spawning-120 hpf	15-250 mg/L	8	292
Wang et al. 2010 [86]	Fruit fly	30, 48, 72, 96, 110 ael (larval stages)	Length[b]	Food	Egg laying-110 ael	100-500 μM	2	378
Wang et al. 2010 [86]	Fruit fly	Pupae	Weight	Food	Egg laying-white pupae stage	100-500 μM	2	98
Pinkas et al. 2010 [187]	Chicken	Hatchling	Weight	Egg injection	Single treatment at incubation day 0	5-10 mg/kg egg	2	52
O'Brien et al. 2009 [236]	Chicken	Embryo at pipping star or day 22, whichever came first	Weight	Egg injection	Single treatment at incubation day 0	0.01-10 mg/kg egg	4[c]	37
Jiang et al. 2012 [3926]	Chicken	Embryonic day 19	Yolk free body weight	Egg injection	Single treatment at incubation day 0	0.5-2 mg/kg egg	2[c]	40
Jiang et al. 2012 [3926]	Chicken	16-24 hours post hatching	Yolk free body weight and crown to rump length	Egg injection	Single treatment at incubation day 0	0.5-2 mg/kg egg	2[c]	68
Spachmo and Arukwe 2012 [3932]	Salmon	Study days 21, 35, 49, 56 (larval stages post hatching)	Length and dry weight	Egg immersion	Egg stage-day 48	100 μg/L	1	80

(hpf) = hours post fertilization. (ael) = hours after egg laying.

[a]Excludes control groups; 1 control group unless otherwise specified. [b]Length measurements provided by study author (used to calculate volume outcome reported in study). [c]2 control groups.

Table 5. Mammalian summary of findings, quality of evidence, and strength of evidence.

Factor	Rating	Basis
Risk of bias across studies	-1	'Allocation concealment' and 'blinding' risks of bias were: (1) truly present; and (2) these risks of bias are shown empirically to influence study outcome in preclinical experimental animal studies.
Indirectness	0	Mammalian data are empirically recognized as direct evidence of human health (Kimmel et al. 1984; US EPA 1996) and there are no data to counteract this assumption.
Inconsistency	0	Point estimates across similar studies (e.g. mouse gavage) are consistent with overlapping confidence bounds. Estimates of change in birth weight from studies in meta-analysis are consistently in the same direction and have low heterogeneity. Results are also consistent in magnitude and direction of effect estimates. Results of the meta-analysis do not appear to be strongly influenced by an individual study.
Imprecision	0	Mammalian data included in meta-analysis showed relatively small confidence intervals in final estimates. Although some studies don't report confidence intervals, data show statistically significant responses at high doses—indicating small confidence intervals.
Publication bias	0	We found no reason to suspect publication bias. The studies were consistent among their findings regardless of size and funding source; the search was comprehensive, and no unpublished studies were found that presented results out of the range of estimates reported by published studies.
Overall quality of evidence (initial rating is 'High')	Moderate	'High' + (- 1) = 'Moderate'
Summary of findings from meta-analysis	NA	Average change in birth weight = -0.023g [-0.029, -0.016] per 1-unit increase in dose (mg/kg BW/day)
Summary of findings from qualitative analysis	NA	The dose-response data showed mixed results, generally with lower doses showing increased weight compared to the control group (mostly non-significant) and higher doses showing decreased weight (both statistically significant and not).
Overall strength of evidence	Sufficient	

Studies included in meta-analysis [source ID]: Abbott et al. 2007 [528], Hines et al. 2009 [260], Lau et al. 2006 [635] (birth weight data), White et al. 2007 [566], White et al. 2009 [312], White et al. 2011 [3862], Wolf et al. 2007 [571] (cross foster and windows of sensitivity data).

Other studies [source ID]: Boberg 2008 et al. [3061], Fenton et al. 2009 [264], Hinderliter et al. 2005 [711], Hu et al. 2010 [68], Lau et al. 2006 [635] (fetal weight data), Onishchenko et al. 2011 [3610], Staples 1984 et al. [1871], Yahia et al. 2010 [103], York 2002 [5122].

-1 = 1 level downgrade in quality.

0 = no change in quality.

Table 6. Non-mammalian summary of findings, quality of evidence, and strength of evidence.

Factor	Rating	Basis
Risk of bias across studies	-1	'Sequence generation', 'allocation concealment' and 'blinding' risks of bias were: (1) truly present; and (2) these risks of bias are shown to empirically matter to study outcome in preclinical experimental animal studies.
Indirectness	-1	We lacked an empirical basis to support that these non-mammalian data were directly relevant to the human health outcome of interest and the routes of exposure varied from how humans would be exposed to PFOA. Evidence that support indirectness are: embryonic development in mammalian organisms (i.e., *in utero* development and live birth) is fundamentally different from development in non-mammalian organisms (i.e., development in egg and hatching); and the route of exposures for the non-mammalian organisms (i.e., eggs injected with or immersed in PFOA-containing solution) are not applicable to humans or other mammalian organisms.
Inconsistency	0	Results appear to divide based on measurement of outcome (weight vs. length); however results are consistent between comparable studies (comparable for outcome, species, and exposure route).
Imprecision	0	The zebrafish and fruit fly data have a relatively large sample size and while no confidence bounds are given, the effect estimates are reasonably close to each other (-5 to -20 percent change). Although some studies don't report confidence intervals, data show statistically significant responses at high doses—indicating small confidence intervals.
Publication bias	0	We found no reason to suspect publication bias. The search was comprehensive, the studies of various sizes and funding sources and no unpublished studies were found that presented results out of the range of estimates reported by published studies.
Overall quality of evidence (initial rating is 'High')	Low	'High' + (- 2) = 'Low'
Summary of findings from qualitative analysis	NA	Dose-response data show mostly non-statistically significant increases in body weight, even at the highest tested doses. The length data show mixed results, with two studies demonstrating statistically significant decreases in length and the other two studies showing statistically non-significant increases in length.

Studies [source ID]: Hagenaars et al. 2011 [59], Jiang et al. 2012 [3926], O'Brien et al. 2009 [236], Pinkas et al. 2010 [187], Spachmo and Arukwe 2012 [3932], Wang et al. 2010 [86].

-1 = 1 level downgrade in quality.

0 = no change in quality.

Figure legends

Figure 1. Flowchart for evaluating risk of bias, quality of evidence, and strength of evidence.

Figure 2. Flowchart of the study selection process.

Figure 3. Risk of bias graphs. Review authors' judgments (*low*, *probably low, probably high and high* risk of bias) about (A) each risk of bias item for each included study and (B) as percentages across all included studies , separated into mammalian and non-mammalian groups. A total of 15 mammalian and 6 non-mammalian studies were included in the review.

Figure 4. Meta-analysis results from combining relevant mouse studies where dams were treated with PFOA via gavage and progeny weight was measured at or soon after birth. Meta-analysis results are from a two-step mixed effects model. Mean [95% confidence interval] change in body weight (g) per 1-unit increase in dose (mg/kg BW/day) is presented. Error bars indicate 95% confidence intervals. Each box represents the dose-response slope estimate for a study; the mid-point of the box is the slope estimated for that study and the box area is proportional to the weight given to each study in the meta-analysis. The diamond is centered at overall meta-analysis slope estimate. Wolf 2007 study split into 2 datasets; a) cross foster (exposure GD1-17); b) windows of sensitivity (exposure groups GD7-17, GD10-17, GD13-17, GD15-17). RE=random effects.

Figure 5. Combined scatter plots of response for each tested dose of PFOA, for all included mammalian studies. Response was measured as the percentage weight change for progeny (A) near-term, and/or (B) at birth. Different colors represent different studies (separated by dotted lines) and different symbols represent different species/exposure route categories. Multiple dots

of the same color represent responses at multiple tested doses within the same study. For each study, doses decrease as y-axis increases and are scaled appropriately (i.e., larger vertical gaps indicate larger gaps between doses), and the minimum dose for all studies is zero. *(within symbols) = statistically significant ($p<0.05$) difference in response compared to control group. ** = mg/kg BW/day, unless otherwise specified. [#] = study split into 2 datasets; a) cross foster (exposure GD1-17); b) windows of sensitivity (exposure groups GD7-17, GD10-17, GD13-17, GD15-17). 95% confidence intervals for the point estimates shown in the figure are provided in Supplemental Material, Tables S46 and S47.

Figure 6. Combined scatter plots of response for each tested dose of PFOA, for all included non-mammalian studies. Response was measured as the percentage (A) weight change, and/or (B) length change. Different colors represent different studies (separated by dotted lines) and different symbols represent different species/exposure route categories. Multiple dots of the same color represent responses at multiple tested doses within the same study. For each study, doses decrease as y-axis increases and are scaled appropriately (i.e., larger vertical gaps indicate larger gaps between doses), and the minimum dose for all studies is zero. *(within symbols) = statistically significant ($p<0.05$) difference in response compared to control group. ** = study split into 2 datasets based on time of outcome measurement a) embryonic day 19; b) 16-24 hours post hatching. 95% confidence intervals for the point estimates shown in the figure are provided in Supplemental Material, Tables S48 and S49.

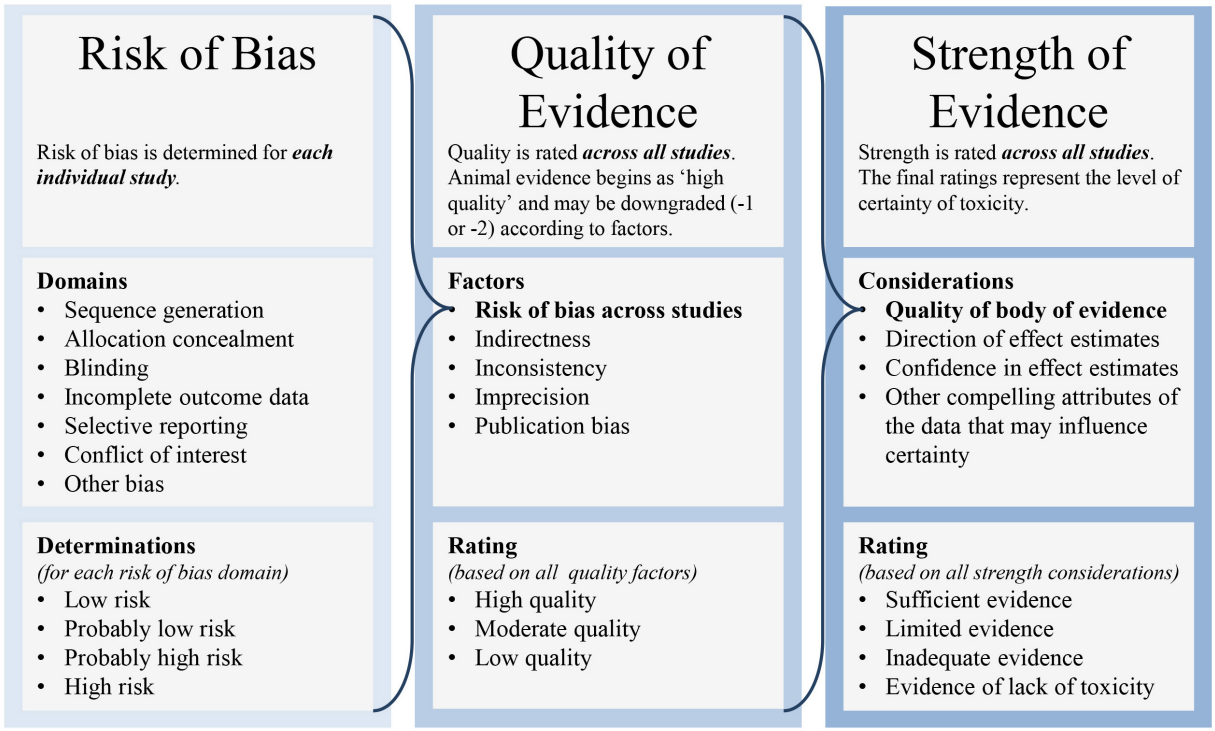

Risk of Bias
Risk of bias is determined for each individual study.
Domains
Sequence generation
Allocation concealment
Blinding
Incomplete outcome data
Selective reporting
Conflict of interest
Other bias
Determinations
(for each risk of bias domain)
Low risk
Probably low risk
Probably high risk
High risk
Quality of Evidence
Quality is rated across all studies. Animal evidence begins as 'high quality' and may be downgraded (-1 or -2) according to factors.
Factors
Risk of bias across studies
Indirectness
Inconsistency
Imprecision
Publication bias
Rating
(based on all quality factors)
High quality
Moderate quality
Low quality
Strength of Evidence
Strength is rated across all studies. The final ratings represent the level of certainty of toxicity.
Considerations
Quality of body of evidence
Direction of effect estimates
Confidence in effect estimates
Other compelling attributes of the data that may influence certainty
Rating
(based on all strength considerations)
Sufficient evidence
Limited evidence
Inadequate evidence
Evidence of lack of toxicity

Figure 1

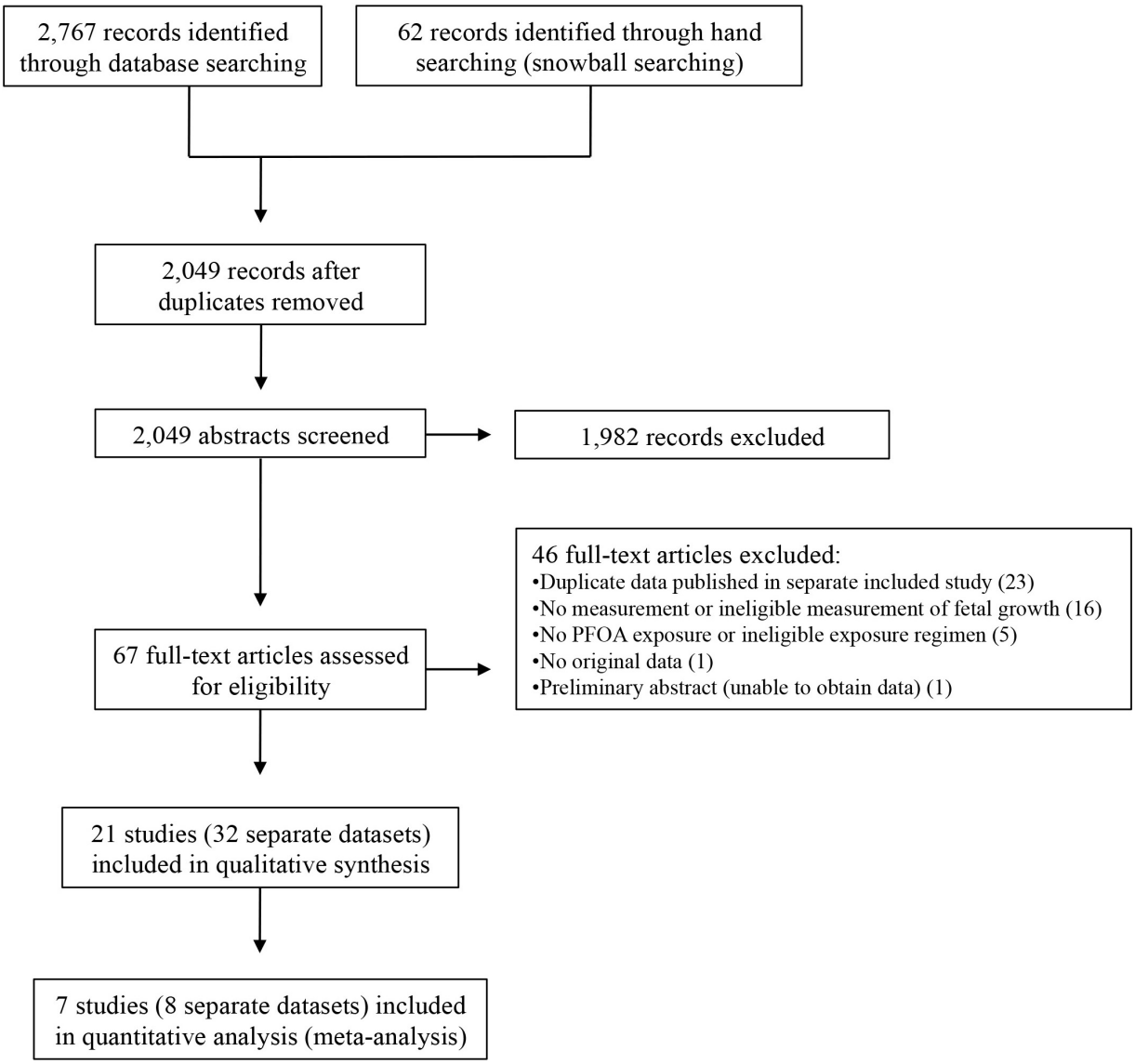

2,767 records identified through database searching
62 records identified through hand searching (snowball searching)
2,049 records after duplicates removed
2,049 abstracts screened
1,982 records excluded
46 full-text articles excluded:
•Duplicate data published in separate included study (23)
•No measurement or ineligible measurement of fetal growth (16)
•No PFOA exposure or ineligible exposure regimen (5)
•No original data (1)
•Preliminary abstract (unable to obtain data) (1)
67 full-text articles assessed for eligibility
21 studies (32 separate datasets) included in qualitative synthesis
7 studies (8 separate datasets) included in quantitative analysis (meta-analysis)

Figure 2

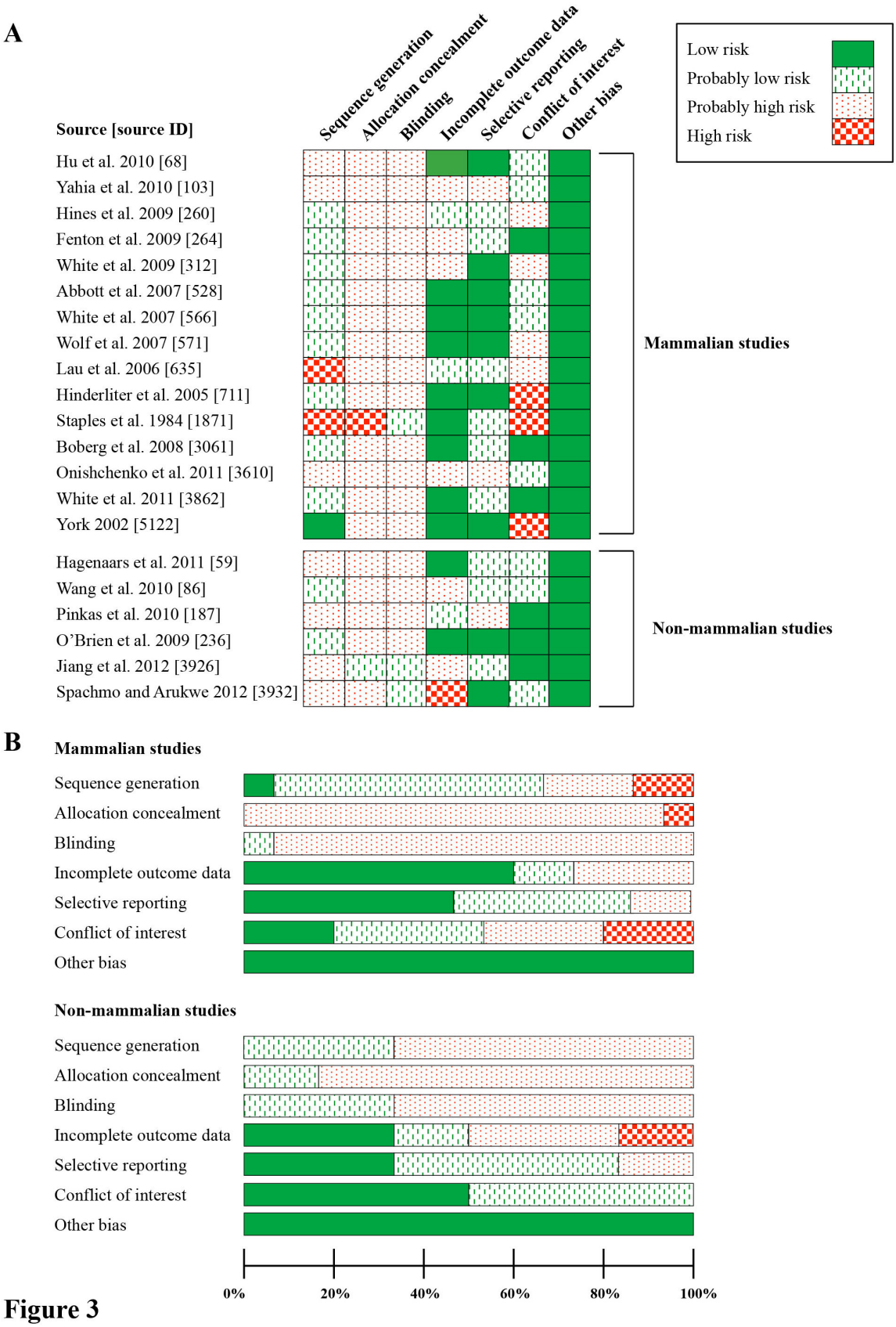

A
Sequence generation
Allocation concealment
Blinding
Incomplete outcome data
Selective reporting
Conflict of interest
Other bias
Low risk
Probably low risk
Probably high risk
High risk
Source [source ID]
Hu et al. 2010 [68]
Yahia et al. 2010 [103]
Hines et al. 2009 [260]
Fenton et al. 2009 [264]
White et al. 2009 [312]
Abbott et al. 2007 [528]
White et al. 2007 [566]
Wolf et al. 2007 [571]
Lau et al. 2006 [635]
Hinderliter et al. 2005 [711]
Staples et al. 1984 [1871]
Boberg et al. 2008 [3061]
Onishchenko et al. 2011 [3610]
White et al. 2011 [3862]
York 2002 [5122]
Mammalian studies
Hagenaars et al. 2011 [59]
Wang et al. 2010 [86]
Pinkas et al. 2010 [187]
O'Brien et al. 2009 [236]
Jiang et al. 2012 [3926]
Spachmo and Arukwe 2012 [3932]
Non-mammalian studies
B
Mammalian studies
Sequence generation
Allocation concealment
Blinding
Incomplete outcome data
Selective reporting
Conflict of interest
Other bias
Non-mammalian studies
Sequence generation
Allocation concealment
Blinding
Incomplete outcome data
Selective reporting
Conflict of interest
Other bias
0%
20%
40%
60%
80%
100%

Figure 3

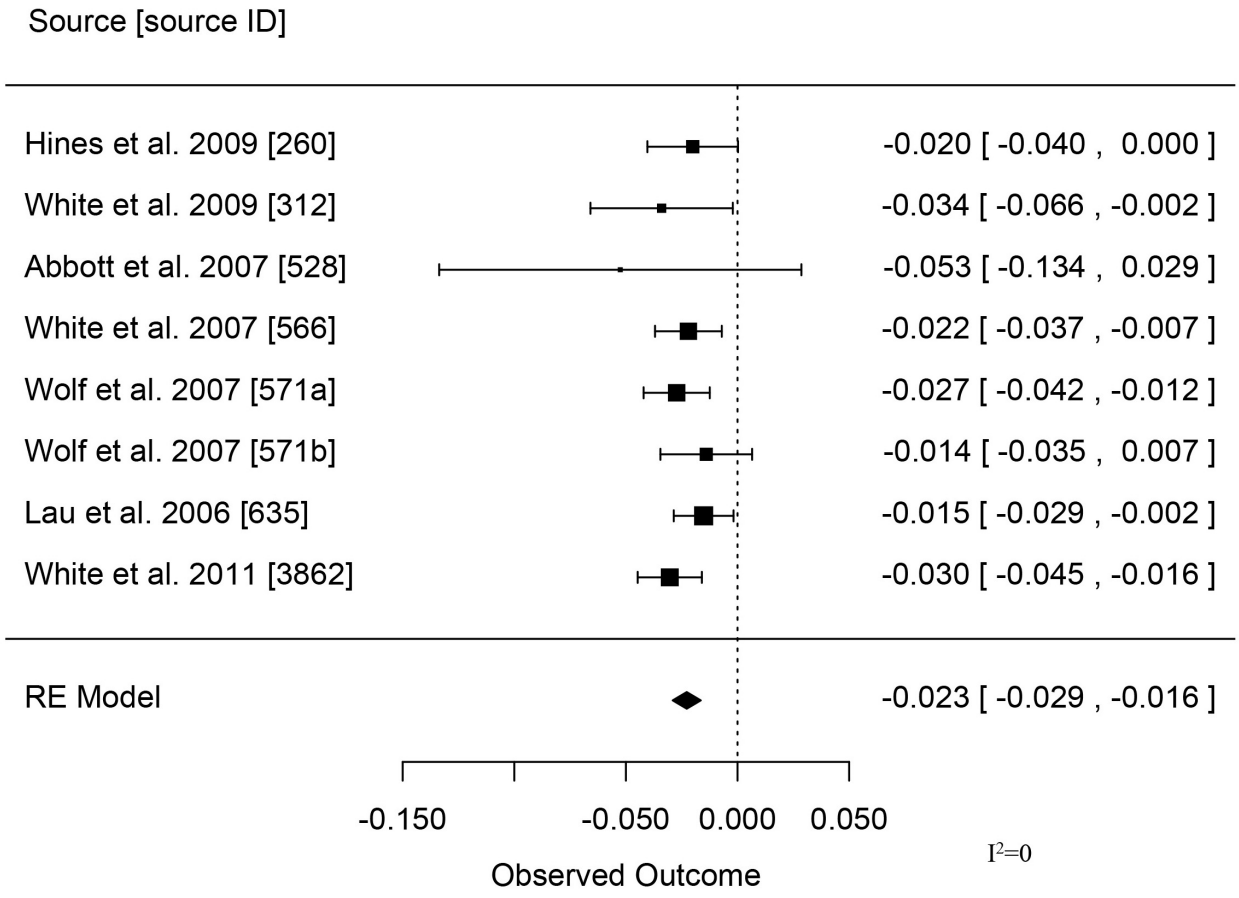
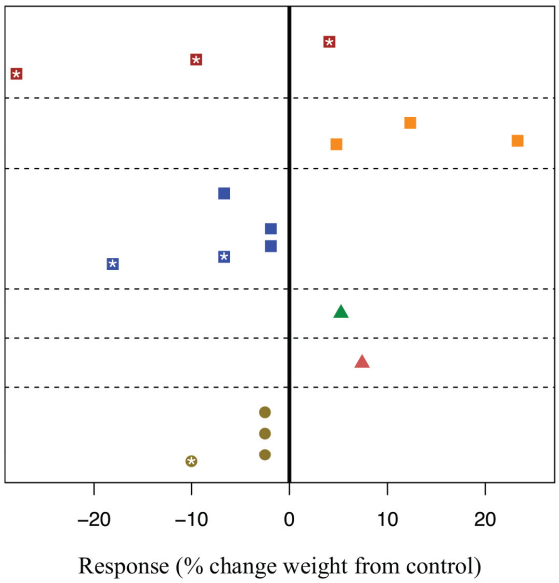

Source [source ID]
Hines et al. 2009 [260]
-0.020 [-0.040 , 0.000]
White et al. 2009 [312]
-0.034 [-0.066 , -0.002]
Abbott et al. 2007 [528]
-0.053 [-0.134 , 0.029]
White et al. 2007 [566]
-0.022 [-0.037 , -0.007]
Wolf et al. 2007 [571a]
-0.027 [-0.042 , -0.012]
Wolf et al. 2007 [571b]
-0.014 [-0.035 , 0.007]
Lau et al. 2006 [635]
-0.015 [-0.029 , -0.002]
White et al. 2011 [3862]
-0.030 [-0.045 , -0.016]
RE Model
-0.023 [-0.029 , -0.016]
-0.150
-0.050
0.000
0.050
Observed Outcome
I²=0

Figure 4

A

Source [source ID]	Species	Route of exposure	Maximum dose**
Yahia et al. 2010 [103]	Mouse	Gavage	10
Fenton et al. 2009 [264]			5
Lau et al. 2006 [635]			40
Staples et al. 1984 [1871]	Rat	Gavage	100
Boberg et al. 2008 [3061]			20
Staples et al.1984 [1871]	Rat	Inhalation	25 mg/m³



B

Source [source ID]	Species	Route of exposure	Maximum dose**
Hu et al. 2010 [68]	Mouse	Drinking Water	1
Onishchenko et al. 2011 [3610]	Mouse	Food	0.3
Yahia et al. 2010 [103]	Mouse	Gavage	10
Hines et al. 2009 [260]			5
Fenton et al. 2009 [264]			5
White et al. 2009 [312]			5
Abbott et al. 2007 [528]			1
White et al. 2007 [566]			5
Wolf et al. 2007 [571b]#			20
Wolf et al. 2007 [571a]#			5
Lau et al. 2006 [635]			20
White et al. 2011 [3862]			5
Hinderliter et al. 2005 [711]	Rat	Gavage	30
Staples et al. 1984 [1871]			100
York 2002 [5122]			30
Staples et al. 1984 [1871]	Rat	Inhalation	25 mg/m³

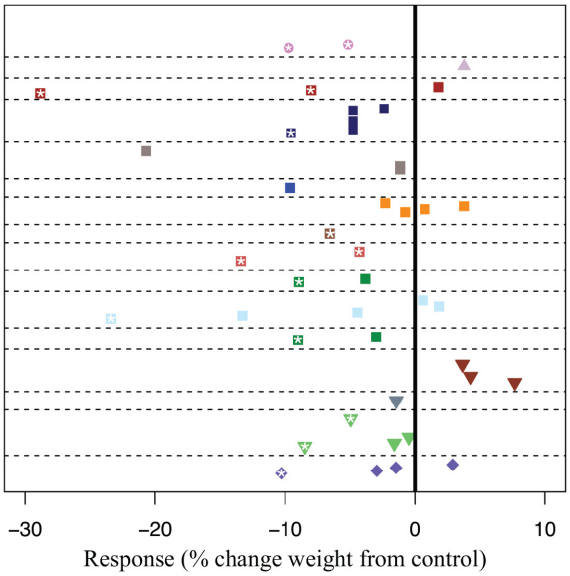
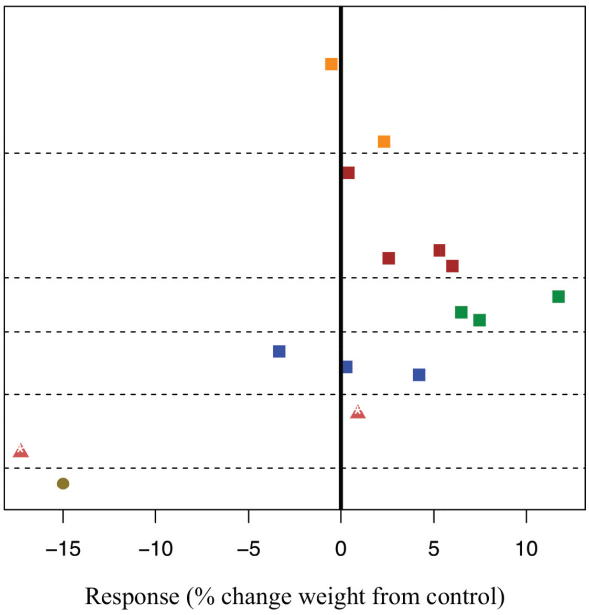


Figure 5

A

Source [source ID]	Species	Route of exposure	Maximum dose
Pinkas et al. 2010 [187]	Chicken	Egg injection	10 mg/kg egg
O'Brien et al. 2009 [236]			10 mg/kg egg
Jiang et al. 2012 [3926a]**			2 mg/kg egg
Jiang et al. 2012 [3926b]**			2 mg/kg egg
Wang et al. 2010 [86]	Fruit fly	Food	500μM
Spachmo and Arukwe 2012 [3932]	Salmon	Egg immersion	100μg/L water



B

Source [source ID]	Species	Route of exposure	Maximum dose
Jiang et al. 2012 [3926]	Chicken	Egg injection	2 mg/kg egg
Wang et al. 2010 [86]	Fruit fly	Food	500μM
Spachmo and Arukwe 2012 [3932]	Salmon	Egg immersion	100μg/L water
Hagenaars et al. 2011 [59]	Zebrafish	Egg immersion	250mg/L water

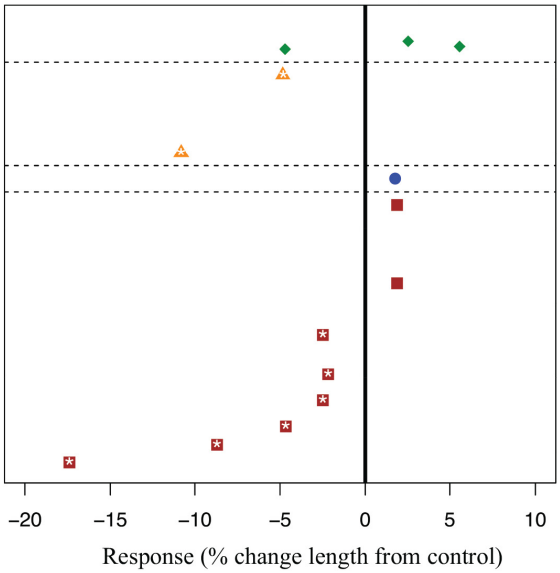


Figure 6